

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



Grado en Tecnologías y Servicios de
Telecomunicación

TRABAJO FIN DE GRADO

INTERFAZ GRÁFICA DE ETIQUETADO DE ATRIBUTOS FACIALES

Autor: Beatriz Cid Fernández

Tutor: Ester González Sosa

Ponente: Julián Fierrez Aguilar

Mayo 2016

INTERFAZ GRÁFICA DE ETIQUETADO DE ATRIBUTOS FACIALES

Autor: Beatriz Cid Fernández
Tutor: Ester González Sosa
Ponente: Julián Fierrez Aguilar



Biometric Recognition Group - ATVS
Dpto. de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Mayo 2016

Resumen

Los sistemas tradicionales de identificación presentan algunas carencias, pues requieren de información como claves de acceso, documentos de identidad o tarjetas de acceso que se pueden perder, robar o incluso falsificar fácilmente. Es por ello que surge el reconocimiento biométrico, cuya finalidad consiste en identificar al individuo a través de información característica propia como puede ser la cara, huella digital, etc. Dichas características propias son conocidas como Soft Biometrics. Se tratan de características humanas físicas (color de ojos, color de pelo, presencia de barba, etc), conductuales (dinámica de tecleo, forma de escribir, forma de andar, etc) y adheridas a características humanas (color de ropa, presencia de tatuajes, accesorios, etc), establecidas y probadas en el tiempo por los seres humanos con el objetivo de diferenciar a los individuos. Estas características físicas y de comportamiento humano se utilizan cada vez más en aplicaciones de seguridad debido a diversas ventajas, tales como la universalidad, la robustez, la permanencia y la accesibilidad.

En el presente Trabajo Fin de Grado, se estudia, implementa y evalúa un sistema de reconocimiento basado en Soft Biometrics. Dichos Soft Biometrics resultan interesantes ya que pueden proporcionar información de un individuo convirtiendo a un sistema biométrico primario más fiable. Para llevarlo a cabo, se desarrolla una interfaz de etiquetado facial, con el fin de estudiar y evaluar cómo afectan un conjunto de Soft Biometrics en la identificación del individuo.

Como punto de partida se ha estudiado el estado del arte en Soft Biometrics y su evolución a lo largo de la historia. Posteriormente se ha diseñado la interfaz gráfica para etiquetado de atributos faciales y se ha procedido al etiquetado de la base de datos LFW, que no dispone de un etiquetado de Soft Biometrics de los individuos, públicamente disponible para investigación. Una vez obtenido el etiquetado, se ha realizado una cuantificación de los resultados obtenidos con el fin de poder operar y extraer conclusiones con dichos datos.

En la parte de análisis, se ha estudiado y evaluado la correlación que existe entre los diferentes Soft Biometrics que se han definido para este proyecto.

En la parte experimental, se ha diseñado un sistema basado en un conjunto de Soft Biometrics en el que se ha estudiado el rendimiento individual y varios conjuntos de ellos atendiendo a distintos criterios.

Por último, se presentan las conclusiones, extraídas a lo largo del trabajo y se proponen líneas de trabajo futuro.

Palabras Clave

reconocimiento biométrico, características, género, edad, raza, posición, base de datos, identificación, clasificación, etiquetado.

Abstract

Traditional identification systems present some limitations, since they require information such as passwords, IDs or access cards which can get lost, stolen or even easily falsified. Thus, biometric recognition arose with the purpose of identifying individuals across their own characteristics such as the face, fingerprint, etc. These characteristics are known as Biometric traits. They involve physical human characteristics (i.e. eyes color, hair color, the presence of beard, etc.), behavioral characteristics (i.e. typing dynamics, writing style, gait, etc.) and characteristics adhered to human characteristics (i.e. clothes color, the presence of tattoos, accessories, etc.), established and proved in the time by human beings with the objective of differentiating individuals. Physical characteristics and human behavior are increasingly being used within security applications due to their several advantages, such as universality, robustness, permanency and accessibility.

The present TFG studies, implements and evaluates a recognition system based on Soft Biometrics. Soft Biometrics are interesting since they may provide information of an individual turning into a reliable primary biometric system. In order to carry it out, a facial labelling interface is developed with the objective of studying and evaluating how a set of Soft Biometrics may perform in identification tasks of an individual.

As starting point, the state of the art in Soft Biometrics and its evolution along history has been studied. Later, the interface for facial attributes labeling has been designed, which has no Soft Biometrics labeling of individuals, labeled. This database was provided by the group of ATVS biometric recognition. As soon as the labeling was obtained, a quantification of the results has been realized with the purpose of operating and extracting conclusions from this data.

In the analysis section, the existing correlation among different Soft Biometrics included in this project has been studied and evaluated.

In the experimental part, a system based on a set of Soft Biometrics has been designed and the individual yield has been studied attending on different criteria.

Finally, the conclusions extracted along the development of the project are presented and future lines of work are proposed.

Key words

Soft Biometrics, biometric recognition, features, gender, age, race, pose, database, identification, classification, labeled.

Agradecimientos

En primer lugar, gracias a Ester, mi tutora, por brindarme la oportunidad de trabajar en este proyecto, por su persistencia, su paciencia y su motivación en todo el proceso y desarrollo del mismo.

Quiero agradecer de igual forma a todos aquellos profesores y compañeros que he tenido la suerte de conocer a lo largo de la carrera, sin los cuales este trabajo tampoco habría sido posible. Gracias a Dani, Mario y Raúl, por todo el apoyo recibido a lo largo de estos años, por hacer las horas de estrés en la biblioteca, antes de un examen, más llevaderas. Sin embargo, mención aparte se merece mi compañera de fatigas, Beatriz, la que empezó siendo una compañera y la que hoy en día se ha convertido en una gran amiga. La que ha estado, está y sé que estará tanto en los buenos momentos como en los malos. Gracias por tu amistad.

A Dani, por su cariño y apoyo diario, por su constante paciencia y por entenderme cada vez que daba prioridad a mis estudios.

Por último y sin duda los más importantes, mis padres, mi hermano y mis abuelos, a los que les doy las gracias por encima de todo. Aquellos que cada vez que hacía un examen esperaban ansiosos mi llegada a casa para preguntarme que tal me había salido. Gracias por todo vuestro amor, esfuerzo y sacrificio sin el cual no estaría donde hoy estoy. Gracias por confiar en mí cuando ni siquiera yo lo hacía y por haber hecho de mí la persona que soy. Gracias por proporcionarme la mejor educación y lecciones de vida. Por enseñarme que con esfuerzo se consigue todo lo que uno se propone.

Muchas gracias a todos.

Índice general

Índice de Figuras	ix
Índice de Tablas	x
1. Introducción	1
1.1. Motivación del proyecto	1
1.2. Objetivos y enfoque	2
1.3. Metodología y plan de trabajo	2
2. Estado del arte. Soft Biometrics	5
2.1. Introducción al Reconocimiento Biométrico	5
2.2. ¿Qué son Soft Biometrics?	7
2.3. Tipos de Soft Biometrics	8
2.4. Aplicaciones	8
2.4.1. Fusión con un rasgo biométrico clásico y reducción de la población de búsqueda . .	9
2.5. Etiquetado de Base de Datos	10
2.5.1. Crowdsourcing y algoritmos automáticos	10
3. Base de Datos LFW	11
3.1. Descripción de la base de datos	11
3.2. Framework y Evaluación LFW	13
4. Diseño de Interfaz Gráfica	15
4.1. Descripción	15
4.1.1. Soft Biometrics elegidos	15
4.2. Diseño de la interfaz	16
4.2.1. Diagrama de Flujo	17
4.2.2. Cuantificación	19
4.3. Explicación de los atributos y criterio de etiquetado	20
5. Análisis estadístico de Soft Biometrics de LFW	23
5.1. Distribución de atributos	23
5.2. Correlación de atributos	27

6. Experimentos realizados y resultados	29
6.1. Medidas de rendimiento de sistemas biométricos	29
6.1.1. Basado en Soft Biometrics	30
6.1.2. Basado en Atributos	31
6.1.3. Rendimiento individual	31
6.1.4. Selección de características a través de SFFS	31
7. Fusión con un sistema basado en cara	35
8. Conclusiones y trabajo futuro	37
Glosario de acrónimos	39
Bibliografía	40
A. Manual del programador	43
A.1. Acerca de GUIDE	43
A.2. Partes de GUIDE	45

Índice de Figuras

1.1. Diagrama del plan de trabajo seguido	2
2.1. Clasificación de rasgos biométricos	5
2.2. Esquema de modo de identificación	6
2.3. Esquema de modo de verificación	7
2.4. Ejemplos de Soft Biometrics. Versión reducida de [1].	7
2.5. Integración de Soft Biometrics para mejorar un sistema biométrico primario.	9
2.6. Integración de Soft Biometrics con un sistema biométrico primario para mejorar la eficiencia de búsqueda.	10
3.1. Ejemplos de la base de datos LFW.	12
4.1. Ejemplos de envejecimiento	16
4.2. Interfaz gráfica diseñada	17
4.3. Funcionamiento de la interfaz diseñada	17
4.4. Diagrama de flujo de la interfaz diseñada	18
5.1. Histogramas Género y Frente	23
5.2. Histogramas Raza y Edad	23
5.3. Histogramas Pose y Ojos	24
5.4. Histogramas Gafas y Sonrisa	25
5.5. Histogramas Barba y Bigote	25
5.6. Histograma de Boca	26
5.7. Correlación a nivel de atributos	28
5.8. Correlación a nivel de instancias	28
6.1. Ejemplo FAR, FRR y EER	30
A.1. Ventana de inicio GUI	44
A.2. Ventana de inicio GUI	44

Índice de Tablas

2.1. Tipos de Soft Biometrics. Tabla traducida de [2].	8
4.1. Oclusión bajo estudio. Oclusion (Oclusión), Nose (Nariz), Mouth (Boca), Chin (Barbilla), Left Ear (Oreja izquierda), Right Ear (Oreja derecha).	15
4.2. Soft Biometrics bajo estudio. Gender (Género), Age (Edad), Race (Raza), Forehead (Frente), Mouth (Boca), Eyes (Ojos), Glasses (Gafas), Smiling (Sonrisa), Beard (Barba), Moustache (Bigote), Pose (Posición), Male (Hombre), Female (Mujer), Baby (Bebé), Child (Niño), Youth (Joven), Middle Aged (Media edad), Senior (Señor), White (Blanco), Black (Negro), Asian (Asiático), Indian (India), Other Mixture (Otra raza), Fully Visible (Totalmente visible), Partially Visible (Parcialmente visible), Obstructed (Obstruido), No (No), Open Widely (Ampliamente abierto), Partially Open (Parcialmente abierto), Close (Cerrado), Yes (Si), Open (Abierto), No Glasses (Sin gafas), Eye Wear (Gafas de ver), Frontal (Frontal), Sunglasses (Gafas de sol), Left Side (Lado izquierdo), Right Side (Lado derecho).	16
4.3. Cuantificación de atributos	19
4.4. Cuantificación de oclusión	19
4.5. Criterio de etiquetado de atributos	21
4.6. Criterio de etiquetado de oclusión	22
5.1. Distribución de Soft Biometrics: Género	24
5.2. Distribución de Soft Biometrics: Edad	24
5.3. Distribución de Soft Biometrics: Raza	24
5.4. Distribución de Soft Biometrics: Frente	24
5.5. Distribución de Soft Biometrics: Posición	24
5.6. Distribución de Soft Biometrics: Ojos	25
5.7. Distribución de Soft Biometrics: Gafas	25
5.8. Distribución de Soft Biometrics: Sonrisa	25
5.9. Distribución de Soft Biometrics: Barba	26
5.10. Distribución de Soft Biometrics: Bigote	26
5.11. Distribución de Soft Biometrics: Boca	26
6.1. EER basado en Soft Biometrics	30
6.2. EER basado en atributos	31
6.3. EER para cada atributo	31
6.4. EER de Soft Biometrics y de atributos	31
6.5. Rendimiento del sistema con datos de entrenamiento basado en el algoritmo de extracción de características SFFS	32

6.6. Rendimiento del sistema con los 10 folds basado en el algoritmo de extracción de características SFFS	32
6.7. Rendimiento del sistema ordenando el menor EER utilizando el algoritmo de extracción de características SFFS	32
7.1. Fusión de un sistema basado en cara con todas las características de la interfaz	35
7.2. Fusión de un sistema basado en cara con las características seleccionadas en SFFS de la interfaz	35
A.1. Herramientas GUI Matlab	45
A.2. Otras Herramientas GUI Matlab	45

1

Introducción

1.1. Motivación del proyecto

Las necesidades de mejorar la seguridad en diferentes contextos de la sociedad actual, da lugar al desarrollo de varias técnicas de reconocimiento automático basadas en rasgos biométricos.

Los rasgos biométricos más utilizados se llevan a cabo mediante el estudio de la cara, huellas dactilares, iris, voz, geometría de la mano y firma.

La autenticación mediante verificación biométrica está convirtiéndose en algo cada vez más habitual en los sistemas de seguridad, tanto privados como públicos, en la electrónica de consumo y en las aplicaciones de punto de venta. La biometría es un excelente método de identificación de la persona que se aplica en muchos procesos debido a dos razones fundamentales, la seguridad y la comodidad, puesto que no requieren de información adicional como documentos de identidad, uso de contraseñas, etc, que son fáciles de olvidar o de falsificar.

En algunas aplicaciones, como pueden ser escenarios de videovigilancia, puede ocurrir que la resolución de la imagen no sea lo suficientemente buena para poder identificar a una persona a través de su región facial. En situaciones de oclusión, tampoco tendríamos acceso al rasgo biométrico concreto. Debido a estas limitaciones, surge la idea de Soft Biometrics como información complementaria del individuo que puede ayudar a sistemas de reconocimiento biométrico cuando su rendimiento no es óptimo. Soft Biometrics tienen como ventajas el hecho que pueden ser adquiridos en la distancia (gran parte de ellos), y están alineados con la manera en la que los seres humanos describen a otros seres humanos. La base de datos LFW, de gran relevancia dentro de la comunidad biométrica, no dispone de un etiquetado de Soft Biometrics de los individuos. Por ello se ha dedicado gran parte del tiempo a etiquetar esta base de datos, con el fin de mejorar tanto los futuros algoritmos automáticos de extracción de Soft Biometrics como mejorar la fusión de sistemas multimodales basados en rasgos clásicos y Soft Biometrics.

La problemática actual que motiva a la realización de este trabajo es la necesidad de obtener resultados competentes que permitan diseñar un robusto sistema de identificación de individuos.

En este Trabajo Fin de Grado, lo que se pretende es hacer un estudio de diferentes atributos y Soft Biometrics faciales para ver cuánta y qué calidad de información puede esto proporcionar al reconocimiento humano.

1.2. Objetivos y enfoque

El objetivo principal de este Trabajo Fin de Grado, como ya se ha comentado anteriormente, es el estudio, implementación y evaluación de un sistema basado en Soft Biometrics, que se lleva a cabo mediante el etiquetado de Soft Biometrics de la base de datos LFW, de manera que permita mejorar en un futuro la identificación de un individuo.

Para ello se ha seguido una serie de pautas que se detallarán a continuación:

1. Estudio detallado del estado del arte: Conocer en qué consisten Soft Biometrics, así como las ventajas e inconvenientes de ellos. Aumentar el conocimiento de los distintos tipos de Soft Biometrics y su contribución a un sistema biométrico primario.
2. Creación de interfaz gráfica: Adquirir los conocimientos necesarios para programar una interfaz gráfica, utilizando Matlab, que permite utilizar una amplia gama de funciones de tratamiento de imágenes y cálculo matemático aplicado en este proyecto.
3. Etiquetado de la base de datos: Dedicar gran parte del tiempo al etiquetado de las imágenes de la base de datos a través de la interfaz diseñada.
4. Estudio de los datos y experimentación: Realizar un estudio del etiquetado realizado para obtener resultados de la correlación entre los distintos atributos y así conocer sus dependencias. Posteriormente se procede a la fase de experimentación con el objetivo de estudiar el rendimiento del sistema basado en Soft Biometrics.
5. Evaluación del trabajo y conclusiones
6. Documentación y redacción de la memoria: En la etapa final se procede a la documentación y redacción de la memoria, concluyendo el trabajo realizado.

1.3. Metodología y plan de trabajo

Con el fin de cumplir los objetivos establecidos al principio del proyecto, se ha seguido un plan de trabajo organizado en el tiempo como muestra la Figura 1.1.

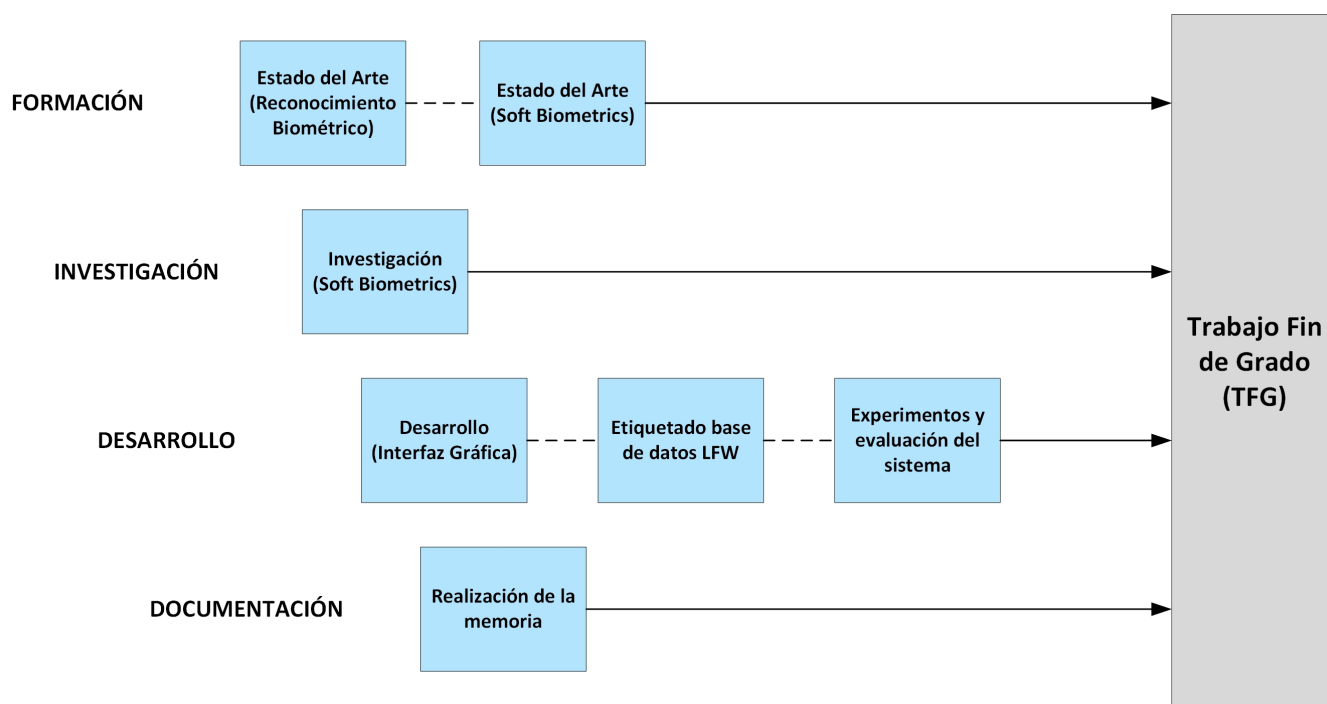


Figura 1.1: Diagrama del plan de trabajo seguido

En primer lugar, se pasa por una etapa de formación en la que se obtienen los conocimientos necesarios para su desarrollo. Por lo tanto, para la realización de este proyecto, se ha estudiado el estado del arte de Soft Biometrics.

Se ha dedicado una gran parte del tiempo de trabajo al desarrollo de una interfaz gráfica, imprescindible para que la tarea del etiquetado de la base de datos sea lo más exacta y rápida posible y con la posibilidad de corregir los errores de una forma rápida y efectiva.

Una vez diseñada dicha interfaz y asegurada su correcta programación y funcionamiento, se ha procedido a realizar el etiquetado de atributos faciales de la base de datos, en el cual se han incluido 12 características bajo estudio.

Posteriormente, se realiza una serie de experimentos de verificación, es decir, dadas dos imágenes y calculados sus Soft Biometrics, decidir si pertenecen a la misma persona o no, y se lleva a cabo un análisis de los datos obtenidos. Esto permitirá, en un futuro, obtener sistemas más robustos para mejorar en la identificación de individuos.

2

Estado del arte. Soft Biometrics

2.1. Introducción al Reconocimiento Biométrico

El reconocimiento biométrico se puede definir como la disciplina que se encarga de establecer la identidad de un individuo basándose en las características físicas, químicas o atributos de comportamiento de la persona. Por lo tanto, se trata de un método automático para el reconocimiento único de humanos, basado en uno o más rasgos conductuales (firma, paso, tecleo) o rasgos físicos intrínsecos (huellas dactilares, retina, iris, patrones faciales, geometría de la mano, etc.) (Véase *Figura 2.1*)

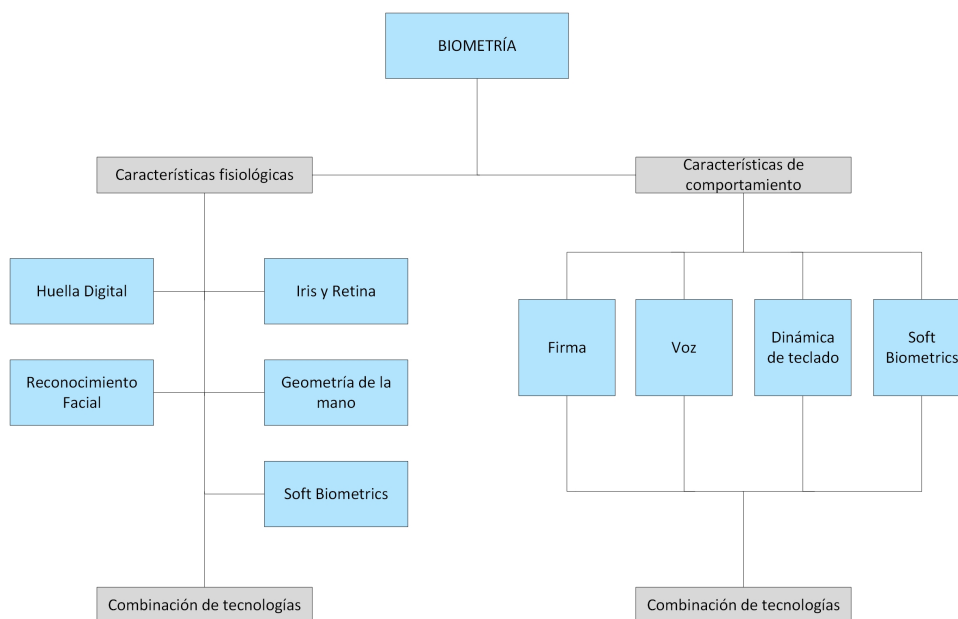


Figura 2.1: Clasificación de rasgos biométricos

Los sistemas biométricos de identificación poseen muchas ventajas sobre los sistemas tradicionales de identificación, pues con estos sistemas no es necesario que el individuo haga uso de contraseñas o tarjetas de identificación que son fáciles de perder, olvidar o robar.

Cualquier característica humana puede ser considerada y aplicada como un rasgo biométrico, siempre y cuando satisfaga los siguientes requerimientos:

- **Universalidad:** El rasgo en cuestión debe estar presente en todos los individuos.
- **Perdurabilidad:** El rasgo debe ser suficientemente invariable en el tiempo.
- **Unicidad:** El rasgo debe permitir discriminar entre individuos.
- **Mensurabilidad:** Se debe poder adquirir el rasgo mediante dispositivos que no causen molestia al individuo.
- **Rendimiento:** Se debe tratar de un sistema preciso. El rendimiento se mide en función de un umbral preestablecido por experimentación, de forma que se pueden obtener dos resultados: aceptación o rechazo. El rendimiento del sistema se mide mediante el EER (*Equal Error Rate*).
- **Aceptabilidad:** Indica el grado de uso de una característica biométrica por parte de los futuros usuarios.
- **Elusión:** Referente a la facilidad con la que el sistema puede ser engañado mediante el uso de métodos fraudulentos.

En el diseño de un sistema biométrico se deben de tener en cuenta varias consideraciones:

- Todo usuario debe ser registrado en el sistema para que su identificador biométrico sea capturado.
- Posteriormente, la información es almacenada en una base de datos central o en una tarjeta inteligente que se le entrega al usuario.
- Finalmente, la información se usará cuando el individuo necesite ser identificado.

Los sistemas de reconocimiento biométrico tienen dos objetivos principales: identificación y verificación del individuo. La identificación consiste en determinar a cuál de todos los usuarios de la base de datos corresponde la muestra de interés, para lo cual se realiza una comparación con todos los modelos de los usuarios registrados. Para ello se hace una comparación 1:N, decidiéndose en función de un umbral cual es el usuario elegido para dicha muestra. La Figura 2.2 muestra el esquema del modo de identificación.

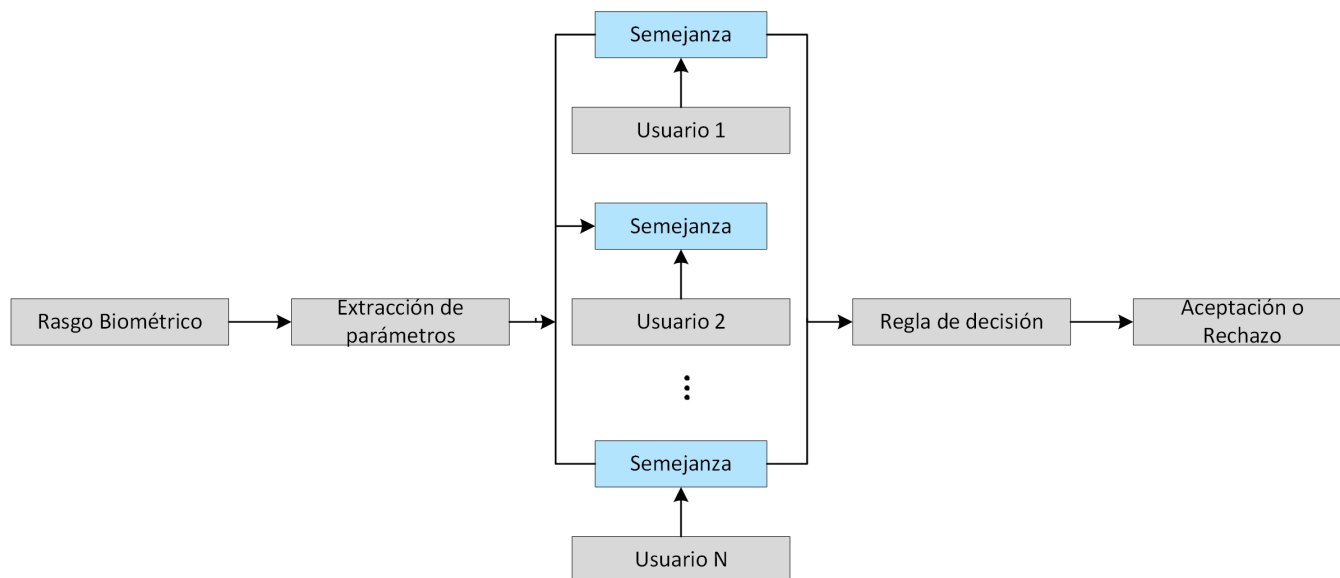


Figura 2.2: Esquema de modo de identificación

La verificación consiste en confirmar o refutar la identidad que algún usuario diga poseer, para ello se aplica una relación 1:1. La decisión de aceptar o rechazar al usuario dependerá de si el valor de parecido supera un determinado umbral. En la Figura 2.3 se puede observar un sistema de modo de verificación.

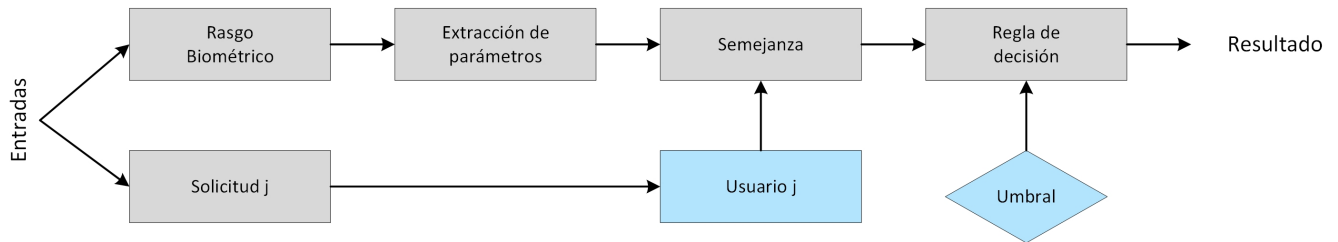


Figura 2.3: Esquema de modo de verificación

2.2. ¿Qué son Soft Biometrics?

Soft Biometric se puede definir como **cualquier característica anatómica o de comportamiento que proporciona alguna información acerca de la identidad de una persona, pero que no presenta pruebas suficientes como para determinar con precisión la identidad de un individuo**. Estas características, al contrario que en la biometría clásica, son establecidas y probadas en el tiempo por los seres humanos con el objetivo de diferenciar a los individuos. Las características personales como el género, raza, edad, altura, peso, color de ojos, cicatrices, marcas, tatuajes o el acento de la voz son ejemplos de Soft Biometrics. Soft Biometrics han ganado cada vez más interés en la biometría por diversas razones, como la necesidad de una mayor precisión y fiabilidad en los sistemas biométricos. El objetivo de Soft Biometrics es aportar información a los identificadores biométricos primarios tales como cara, iris, huellas dactilares y la voz. Por lo tanto, la utilización de Soft Biometrics puede mejorar la precisión en el reconocimiento de los sistemas biométricos primarios, ya sea en escenarios de videovigilancia, presencia de oclusiones, etc.



Figura 2.4: Ejemplos de Soft Biometrics. Versión reducida de [1].

2.3. Tipos de Soft Biometrics

La clasificación de Soft Biometrics es muy amplia y a medida que progrese la tecnología esta clasificación se verá incrementada. Sin embargo, se puede hacer una primera clasificación con los rasgos más dominantes. Soft Biometrics pueden ser continuos o discretos. Rasgos como el género, color de ojos, raza...son variables discretas. Por otro lado, rasgos como la altura o el peso son variables continuas. Un sistema que se basa totalmente en Soft Biometrics, no puede proporcionar la precisión requerida en el reconocimiento de los individuos, por eso están pensados para aportar información añadida a los sistemas biométricos primarios. A medida que el usuario interactúa con el sistema biométrico primario, el sistema debe ser capaz de medir automáticamente Soft Biometrics como la altura, peso, edad, sexo y origen étnico de una manera no intrusiva y sin ninguna interacción con el usuario.

En la Tabla 2.1 podemos observar el tipo de rasgo con su carácter distintivo y el tipo de valor que toma.

Soft Biometric	Cara/Cuerpo/ Accesorio	Naturaleza del Valor	Permanencia	Carácter Distintivo	Percepción Subjetiva
Color de cara	Cara	Continuo	Medio	Bajo	Medio
Color de Pelo	Cara	Continuo	Medio	Medio	Medio
Color de Ojos	Cara	Continuo	Alto	Medio	Medio
Barba	Cara	Binario	Bajo/Medio	Bajo	Medio
Bigote	Cara	Binario	Bajo/Medio	Bajo	Medio
Mediciones Faciales	Cara	Continuo	Alto	Medio	Medio/Alto
Formas Faciales	Cara	Discreto	Alto	Alto	Alto
Mediciones de Características Faciales	Cara	Continuo	Alto	Alto	Medio/Alto
Formas de Características Faciales	Cara	Discreto	Alto	Alto	Alto
Maquillaje	Cara	Discreto	Bajo	Bajo	Medio
Raza	Cara	Discreto	Alto	Medio	Medio
Marcas	Cara/Cuerpo	Discreto	Alto	Medio/Alto	Bajo
Género	Cara/Cuerpo	Binario	Alto	Bajo	Bajo
Edad	Cara/Cuerpo	Continuo	Bajo/Medio	Medio	Medio
Altura	Cuerpo	Continuo	Medio/Alto	Medio	Medio
Peso	Cuerpo	Continuo	Bajo/Medio	Medio	Medio
Paso	Cuerpo	Continuo	Medio	Medio	Alto
Medidas Corporales	Cuerpo	Continuo	Medio/Alto	Medio/Alto	Medio
Formas del Cuerpo	Cuerpo	Discreto	Medio	Medio	Medio
Color de Ropa	Accesorio	Discreto	Bajo	Medio	Medio
Gafas	Accesorio	Binario	Bajo/Medio	Bajo	Bajo

Tabla 2.1: Tipos de Soft Biometrics. Tabla traducida de [2].

2.4. Aplicaciones

La identificación de personas se está convirtiendo en una parte esencial de las necesidades de seguridad en algunas infraestructuras tales como aeropuertos, centros comerciales, edificios gubernamentales y estaciones de tren, entre otros. Aunque estos tipos de rasgos biométricos solo se han considerado recientemente en la biometría, tienen un enorme potencial para la identificación humana, puesto que mejoran el rendimiento del reconocimiento de los rasgos biométricos primarios.

Una de las principales ventajas de Soft Biometrics es su relación con las descripciones humanas convencionales. Los seres humanos, naturalmente hacen uso de Soft Biometrics para identificar y describirse entre ellos. Esto permite la búsqueda de imágenes de vigilancia y bases de datos basadas únicamente en la descripción de un testigo presencial. Por lo tanto la principal ventaja de Soft Biometrics radica en su capacidad para describir a personas desde una perspectiva humana, por lo que se podría decir que

sirve de puente entre la máquina y las descripciones humanas de una persona. Se puede encontrar en [3] alguna búsqueda de personas a través de atributos.

2.4.1. Fusión con un rasgo biométrico clásico y reducción de la población de búsqueda

El sistema de reconocimiento biométrico se divide en dos subsistemas. Un subsistema se llama sistema biométrico primario, el cual se basa en identificadores biométricos tradicionales como la huella digital, cara, geometría de la mano, entre otros. El segundo subsistema hace referencia al sistema biométrico secundario, el cual se basa en Soft Biometrics como la edad, el sexo y la altura. La fusión con el rasgo biométrico clásico es un proceso que se puede hacer o bien en la etapa de extracción de características o bien en la etapa de clasificación. La Figura 2.5 muestra la arquitectura de un sistema de identificación personal que hace uso de ambos subsistemas.

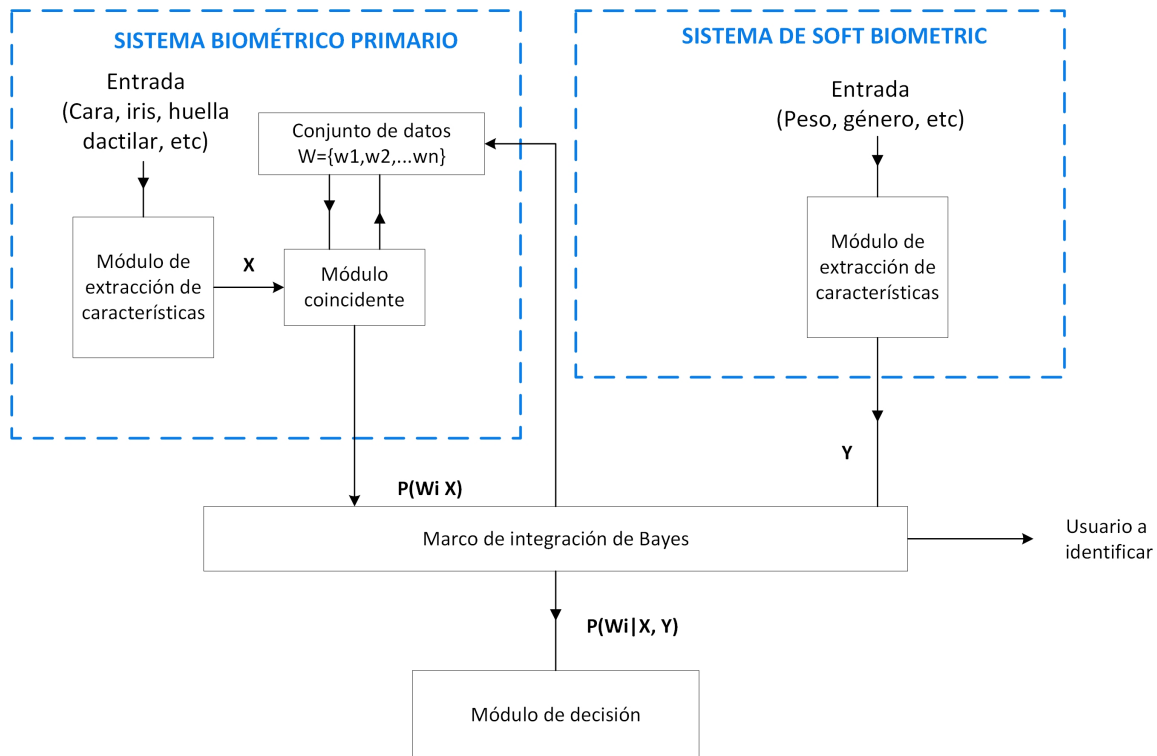


Figura 2.5: Integración de Soft Biometrics para mejorar un sistema biométrico primario.

Dado $W = \{w_1, w_2, \dots, w_n\}$ que representa los n usuarios registrados en la base de datos y X es el vector de características correspondiente a la biometría primaria, se puede asumir que la salida del sistema biométrico primario es de la forma $P(w_i | x)$, $i = 1, 2, \dots, n$, donde $P(w_i | x)$ es la probabilidad de que el usuario de prueba sea w_i dado el vector de características X . Dado $Y = \{y_1, y_2, \dots, y_m\}$, siendo el vector de Soft Biometrics, la probabilidad del usuario w_i , dado el vector x y el vector y , se puede calcular utilizando la regla de Bayes:

$$P(w_i | x, y) = \frac{p(y | w_i) P(w_i | x)}{\sum_{j=1}^n p(y | w_j) P(w_j | x)} \quad (2.1)$$

En la práctica no todos los Soft Biometrics aportan la misma cantidad de información, por lo tanto es necesario incluir un sistema de ponderación de Soft Biometrics basado en un índice de carácter distintivo y permanencia.

Aparte de la mejora del rendimiento de un sistema biométrico primario, también se hace uso de Soft Biometrics para mejorar la eficiencia de búsqueda del sistema biométrico, como se muestra en la Figura 2.6.

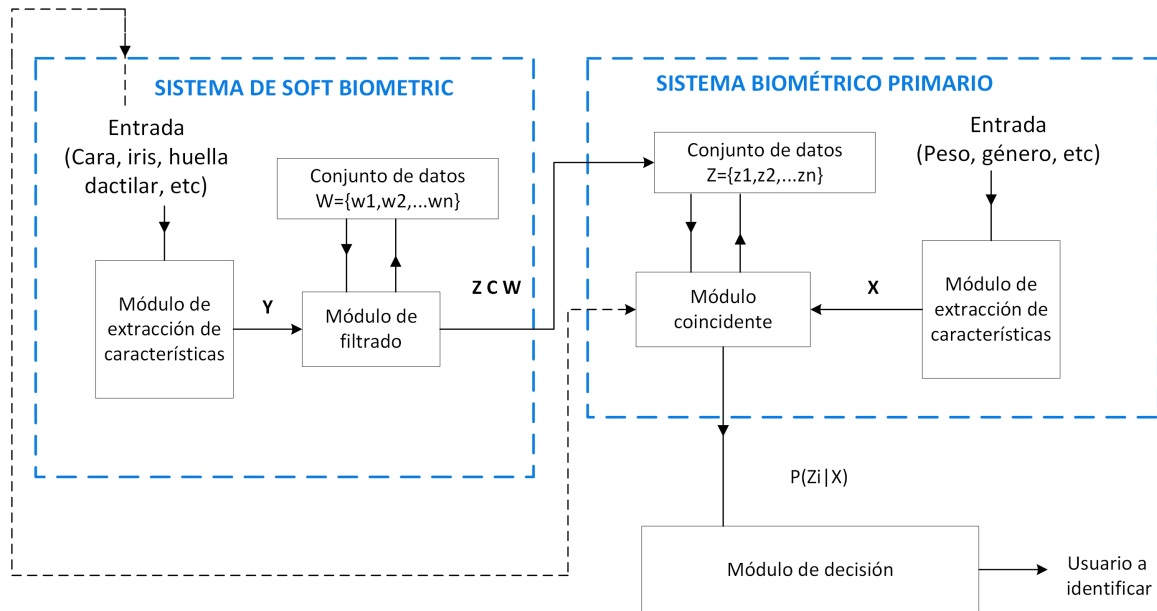


Figura 2.6: Integración de Soft Biometrics con un sistema biométrico primario para mejorar la eficiencia de búsqueda.

Como se observa en la figura, el filtrado se refiere a la limitación del número de entradas en la base de datos que se desea buscar. Por ejemplo, si el usuario se identifica como un varón de mediana edad, la búsqueda puede limitarse únicamente a los individuos inscritos en la base de datos con ese perfil. Esto mejora en gran medida la eficiencia de la búsqueda del sistema biométrico. Pero se debe tener en especial cuidado al diseñar un sistema biométrico con esta particularidad, pues mientras el filtrado reduce el tiempo necesario para la identificación, los errores que se producen en el filtrado pueden degradar el rendimiento del reconocimiento. Wayman [4] propuso como ejemplo el uso del género y la edad para filtrar una gran base de datos biométrica.

2.5. Etiquetado de Base de Datos

2.5.1. Crowdsourcing y algoritmos automáticos

Existen algunas plataformas que permiten realizar el etiquetado automático de una base de datos. Algunos de estos ejemplos son Amazon Mechanical Turk o CrowdFlower. Se puede profundizar más en [5] .

3

Base de Datos LFW

3.1. Descripción de la base de datos

LFW (Labeled Faces in the Wild) es una base de datos de referencia en la comunidad de reconocimiento facial que contiene una serie de fotografías de caras diseñada para estudiar el problema del reconocimiento facial. Se puede encontrar en la página oficial <http://vis-www.cs.umass.edu/lfw/> [6]. Todas estas imágenes han sido recogidas utilizando el detector de caras de Viola-Jones (primer marco de detección de objetos para proporcionar tasas de detección de objetos competitivas en tiempo real) y se lleva a cabo en la biblioteca OpenCV.

- La base de datos contiene 13233 imágenes de caras de personajes recogidos en internet. Algunas imágenes contienen más de una cara, pero es la cara que contiene el pixel central la que se elige, quedando las demás caras ignoradas como fondo.
- Cada imagen viene etiquetada con un nombre único e irrepetible, por lo que un nombre no debe de corresponder a más de una persona.
- La base de datos contiene imágenes de 5749 individuos diferentes, de los cuales, 1680 personas tienen dos o más imágenes en la base de datos. Las restantes 4069 tienen sólo una imagen en la base de datos.
- Las imágenes están disponibles como imágenes JPEG de 250x250 píxeles. La mayoría de las imágenes son en color, sin embargo unas pocas vienen dadas en escala de grises.

En la Figura 3.1 se muestran algunos ejemplos de dicha base de datos:



(a) Pares coincidentes de la base de datos LFW. Estas imágenes muestran una serie de características de dicha base: Una persona puede tener más de una imagen.

(b) Pares no coincidentes de la base de datos LFW

Figura 3.1: Ejemplos de la base de datos LFW.

3.2. Framework y Evaluación LFW

La base de datos LFW, es de suma importancia, pues es una base de datos recogida de forma natural, es decir, sin preparar el ambiente para conseguir una iluminación específica y sin manipular a la persona para que adopte una posición en particular. La mayoría de las bases de datos existentes se han creado bajo condiciones controladas para facilitar el estudio de los parámetros específicos sobre el problema de reconocimiento facial. Estos parámetros incluyen variables tales como la posición, postura, iluminación, expresión, fondo, calidad de la cámara, la oclusión, edad y género. Sin embargo, la base de datos LFW, representa un primer intento de proporcionar un conjunto de fotografías faciales que abarcan el conjunto de condiciones que habitualmente se encuentran en la vida cotidiana. La tarea típica se llama verificación facial: dado un par de imágenes, un clasificador binario debe predecir si las dos imágenes son de la misma persona. Otra tarea típica es la identificación facial: dada la imagen de la cara de una persona desconocida, se debe identificar el nombre de la persona, haciendo referencia a una galería de imágenes vistas anteriormente de las personas identificadas. El reconocimiento facial es el problema de la identificación de un individuo específico, en lugar de limitarse a la detección de la presencia de la cara de un individuo.

4

Diseño de Interfaz Gráfica

4.1. Descripción

4.1.1. Soft Biometrics elegidos

Con el fin de disponer de información de Soft Biometrics de una base de datos de reconocimiento facial de relevancia, que pueda servir tanto para impulsar el desarrollo de algoritmos automáticos de extracción de Soft Biometrics, como para la fusión de Soft Biometrics con rasgos biométricos tradicionales, se ha desarrollado una interfaz gráfica de etiquetado de atributos faciales, que permitirá una mayor rapidez de etiquetado de la base de datos LFW. Además de estos Soft Biometrics, se ha tenido en cuenta la posibilidad de oclusión de alguna parte de la cara bajo estudio (nariz, boca, barbilla, oreja izquierda y oreja derecha), puesto que en la vida real, la oclusión de alguna parte de la cara de una persona es algo que dificulta la búsqueda del individuo en un sistema de reconocimiento facial. Teniendo en cuenta que un Soft Biometric puede tener varias instancias (p.ej: Género es un Soft Biometric que se compone de dos instancias: hombre y mujer) se ha pensado que los Soft Biometrics más adecuados para este estudio son los recogidos en la Tabla 4.1 y en la Tabla 4.2:

Oclusion
Nose
Mouth
Chin
Left Ear
Right Ear

Tabla 4.1: Oclusión bajo estudio. Oclusion (*Oclusión*), Nose (*Nariz*), Mouth (*Boca*), Chin (*Barbilla*), Left Ear (*Oreja izquierda*), Right Ear (*Oreja derecha*).

Gender	Male	Female			
Age	Baby	Child	Youth	Middle Aged	Senior
Race	White	Black	Asian	Indian	Other Mixture
Forehead	Fully Visible	Partially Visible	Obstructed		
Mouth	Open Widely	Partially Open	Close		
Eyes	Open	Partially Open	Close		
Glasses	No Glasses	Eye Wear	Sunglasses		
Smiling	Yes	No			
Beard	Yes	No			
Moustache	Yes	No			
Pose	Frontal	Left Side	Right Side		

Tabla 4.2: Soft Biometrics bajo estudio. Gender (Género), Age (Edad), Race (Raza), Forehead (Frente), Mouth (Boca), Eyes(Ojos), Glasses (Gafas), Smiling (Sonrisa), Beard (Barba), Moustache (Bigote), Pose(Posición), Male (Hombre), Female (Mujer), Baby (Bebé), Child (Niño), Youth (Joven), Middle Aged(Media edad), Senior (Señor), White (Blanco), Black (Negro), Asian (Asiático), Indian (India), Other Mixture (Otra raza), Fully Visible (Totalmente visible), Partially Visible (Parcialmente visible), Obstructed (Obstruido), No (No), Open Widely (Ampliamente abierto), Partially Open (Parcialmente abierto), Close(Cerrado), Yes (Si), Open (Abierto), No Glasses(Sin gafas), Eye Wear (Gafas de ver), Frontal (Frontal), Sunglasses (Gafas de sol), Left Side (Lado izquierdo), Right Side (Lado derecho).

En algunos casos, la edad es un factor crítico y juega un papel muy importante en el reconocimiento facial, por ejemplo, en los casos de la desaparición de niños o la búsqueda en una base de datos que tiene muchos años. El envejecimiento resulta problemático ya que hay muchos rasgos que cambian con el tiempo y esto dificulta la identificación del individuo. Para los sujetos que pertenecen al grupo juvenil (menores de 18 años) los cambios afectan más a la forma mientras que para la gente mayor de 18 años los cambios son más de textura (cambios de peso, firmeza de la piel).



Figura 4.1: Ejemplos de envejecimiento

4.2. Diseño de la interfaz

Dado que el etiquetado de las características nombradas anteriormente es un proceso bastante largo y costoso, se ha creado una interfaz gráfica, con el fin de simplificar el proceso de etiquetado de características. Es por ello que antes de proceder al diseño de la interfaz, se ha dedicado gran parte del tiempo al estudio sobre el funcionamiento de una GUI en Matlab. (Véase Anexo A).

Posteriormente, se ha descargado la base de datos LFW proporcionada por el Grupo ATVS y se ha creado una Gui básica en Matlab, que se ha ido desarrollando progresivamente, hasta conseguir el resultado final de dicha interfaz la cual permitirá la selección de los diferentes atributos característicos de las distintas imágenes de la base de datos. Una vez programado esto, se ha creado un fichero de texto, donde se irán guardando los diferentes nombres de las imágenes y los atributos elegidos para cada imagen con el siguiente formato:

[Nombre] [Gender] [Age] [Race] [Forehead] [Mouth] [Eyes] [Glasses] [Smiling] [Beard] [Moust.] [Pose]

Otro fichero de texto donde se guardarán las oclusiones de cada imagen con el siguiente formato:

[NombreArchivo] [Nose] [Mouth] [Chin] [LeftEar] [RightEar]

y finalmente otro fichero de texto donde se guardarán los valores cuantificados de cada imagen (Se verá posteriormente).

Los ficheros de texto resultantes, constarán con tantas líneas como imágenes hay en la base de datos, es decir, 13233.

El resultado final de la interfaz diseñada se puede ver en la siguiente figura:



Figura 4.2: Interfaz gráfica diseñada

4.2.1. Diagrama de Flujo

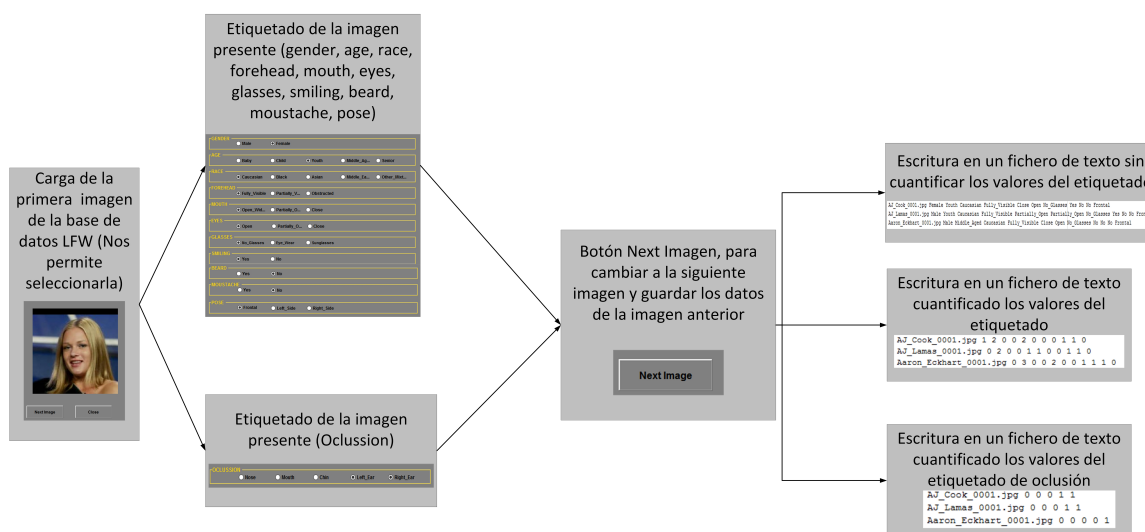


Figura 4.3: Funcionamiento de la interfaz diseñada

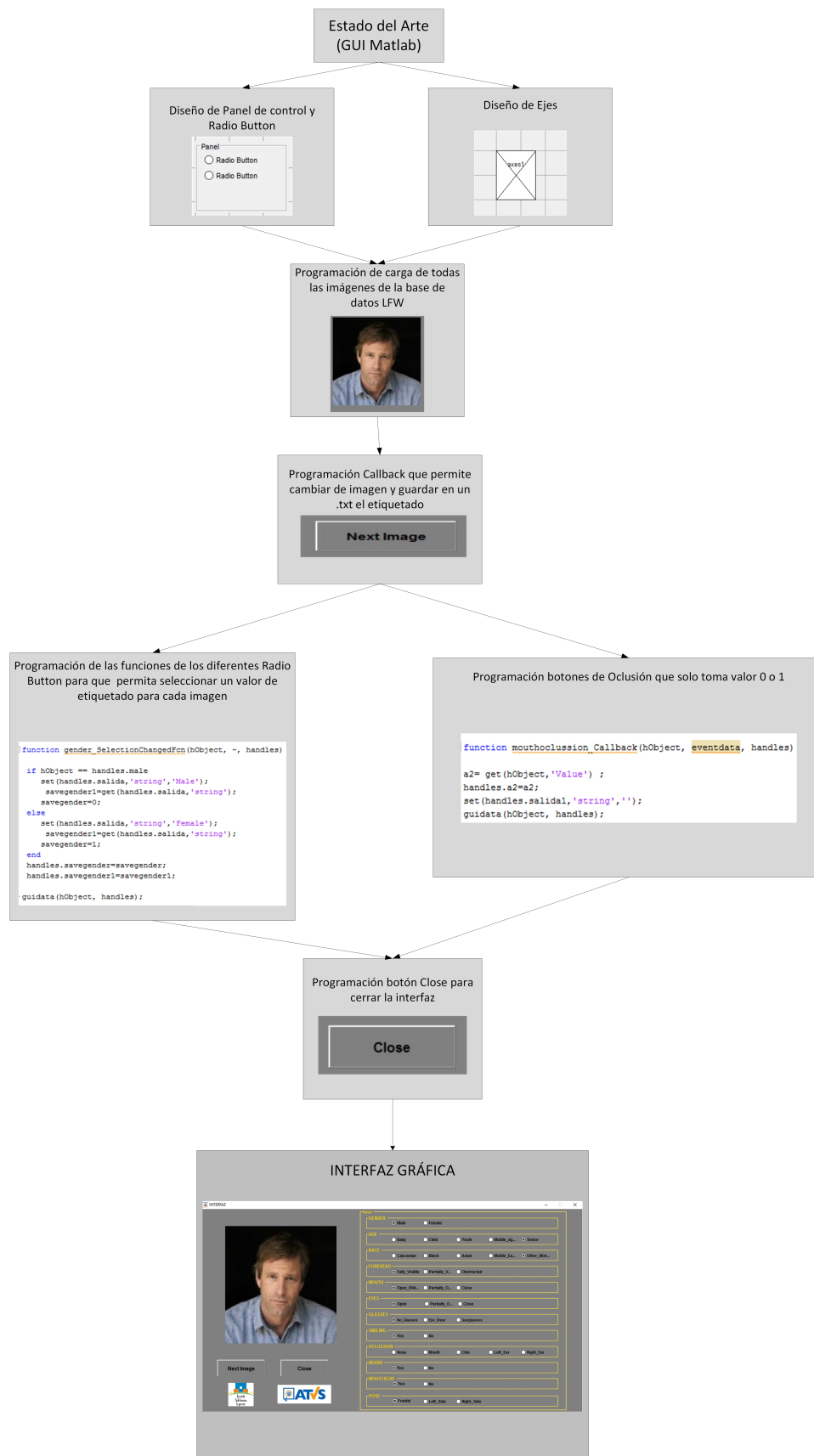


Figura 4.4: Diagrama de flujo de la interfaz diseñada

4.2.2. Cuantificación

De cara a poder trabajar con los resultados obtenidos del etiquetado, ya que con texto no podemos trabajar, se ha procedido a realizar una cuantificación de cada instancia, lo que permitirá finalmente tener un fichero de texto con números que se podrán utilizar para extraer histogramas o la correlación. El criterio de etiquetado que se ha llevado a cabo se puede ver en la tabla 4.2 y en la tabla 4.3.

ATRIBUTO	INSTANCIA	VALOR CUANTIFICADO
GENDER	Male	0
GENDER	Female	1
AGE	Baby	0
AGE	Child	1
AGE	Youth	2
AGE	Middle Aged	3
AGE	Senior	4
RACE	Caucasian	0
RACE	Black	1
RACE	Asian	2
RACE	Middle Eastern	3
RACE	Other Mixture	4
FOREHEAD	Fully Visible	0
FOREHEAD	Partially Visible	1
FOREHEAD	Obstructed	2
MOUTH	Open Widely	0
MOUTH	Partially Open	1
MOUTH	Close	2
EYES	Open	0
EYES	Partially Open	1
EYES	Close	2
GLASSES	No Glasses	0
GLASSES	Eye Wear	1
GLASSES	Sunglasses	2
SMILING	Yes	0
SMILING	No	1
BEARD	Yes	0
BEARD	No	1
MOUSTACHE	Yes	0
MOUSTACHE	No	1
POSE	Frontal	0
POSE	Left Side	1
POSE	Right Side	2

Tabla 4.3: Cuantificación de atributos


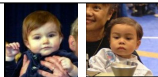
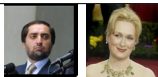







OCLUSIÓN	VALOR CUANTIFICADO
NOSE	0(Si no la tiene oclusa) 1(Si la tiene oclusa)
MOUTH	0(Si no la tiene oclusa) 1(Si la tiene oclusa)
CHIN	0(Si no la tiene oclusa) 1(Si la tiene oclusa)
LEFT EAR	0(Si no la tiene oclusa) 1(Si la tiene oclusa)
RIGHT EAR	0(Si no la tiene oclusa) 1(Si la tiene oclusa)

Tabla 4.4: Cuantificación de oclusión

4.3. Explicación de los atributos y criterio de etiquetado

Antes de comenzar el desarrollo de la interfaz gráfica se han seleccionado una serie de atributos que serán los más adecuados de cara a realizar un estudio facial de los individuos de la base de datos LFW. Es por ello que se han definido un conjunto de 11 atributos: género, edad, raza, frente, boca, ojos, gafas, sonrisa, barba, bigote y posición. Además se ha definido un atributo de oclusión para detectar aquellas personas que tienen alguna zona de la cara no visible lo que dificulta el reconocimiento del individuo.

En la siguiente tabla se nombran los atributos elegidos y el criterio de etiquetado que se ha seguido para conseguir un etiquetado lo más exacto posible, siguiendo siempre las mismas normas:

ATRIBUTO	INSTANCIA	CRITERIO DE ETIQUETADO	EJEMPLOS
GENDER	Male	Si es un hombre.	
GENDER	Female	Si es una mujer.	
AGE	Baby	Edad entre 0 y 4 años.	
AGE	Child	Edad alrededor de 17 años.	
AGE	Youth	Edad entre 18 y 39 años.	
AGE	Middle Aged	Edad entre 40 y 60 años.	
AGE	Senior	Más de 60 años.	
RACE	Caucasian	Si es de piel blanca.	
RACE	Black	Si es de piel negra.	
RACE	Asian	Si tiene los ojos achinados.	
RACE	Middle Eastern	Si tiene rasgos indios.	
RACE	Other Mixture	Si no pertenece a las razas anteriores.	
FOREHEAD	Fully Visible	Si se ve la frente completamente.	
FOREHEAD	Partially Visible	Si se ve la mitad de la frente, ya sea por gorra o flequillo.	
FOREHEAD	Obstructed	Aquella frente que por cualquier razón (pelo u oclusión) está casi o completamente obstruida.	

ATRIBUTO	INSTANCIA	CRITERIO DE ETIQUETADO	EJEMPLOS
MOUTH	Open Widely	Siempre que los dientes de arriba y los de abajo estén separados.	 
MOUTH	Partially Open	Si tiene boca abierta pero los dientes no separados.	 
MOUTH	Close	Si los labios están pegados.	 
EYES	Open	Si tiene los ojos completamente abiertos.	 
EYES	Partially Open	Si tiene los ojos entre abiertos, o uno abierto y el otro cerrado.	 
EYES	Close	Si tiene los ojos cerrados. Si tiene gafas de sol, consideramos que tiene los ojos cerrados.	 
GLASSES	No Glasses	Si no lleva gafas.	 
GLASSES	Eye Wear	Si lleva gafas de ver (con el cristal transparente).	 
GLASSES	Sunglasses	Si lleva gafas de sol (con el cristal oscuro). Si lleva gafas de nieve, consideramos que tiene gafas de sol.	 
SMILING	Yes	Si tiene cara de alegría y se le ven los dientes o si tiene cara de alegría pero no se le ven los dientes.	 
SMILING	No	Si no se le ven los dientes y está serio.	 
BEARD	Yes	Si se ve claramente que tiene barba (Si tiene perilla se considera barba).	 
BEARD	No	Si no se ve claramente que tiene barba.	 
MOUSTACHE	Yes	Si se ve claramente que tiene bigote.	 
MOUSTACHE	No	Si no se ve claramente que tiene bigote.	 
POSE	Frontal	Siempre y cuando los dos ojos sean visibles.	 
POSE	Left Side	Cuando solo sea visible un ojo y siempre que el lado de la cara que se ve sea el derecho (según el punto de vista del observador).	 
POSE	Right Side	Cuando solo sea visible un ojo y, siempre que el lado de la cara que se ve sea el izquierdo (según el punto de vista del observador).	 

Tabla 4.5: Criterio de etiquetado de atributos






OCLUSIÓN	CRITERIO DE ETIQUETADO	EJEMPLOS
NOSE	Si tiene la nariz más tapada que visible.	
MOUTH	Si tiene la boca más tapada que visible.	
CHIN	Si tiene la barbilla más tapada que visible.	
LEFT EAR	Si tiene la oreja izquierda más tapada que visible.	
RIGHT EAR	Si tiene la oreja derecha más tapada que visible.	

Tabla 4.6: Criterio de etiquetado de oclusión

5

Análisis estadístico de Soft Biometrics de LFW

5.1. Distribución de atributos

Una vez realizado el etiquetado completo de la base de datos, se procedió a estudiar su distribución mediante una serie de histogramas, donde se muestran las instancias de cada atributo para tener una idea de qué instancia es la predominante para cada imagen de la base de datos LFW dentro de cada atributo. Debajo de cada histograma se ha representado una tabla en la que se muestran los porcentajes para cada instancia:

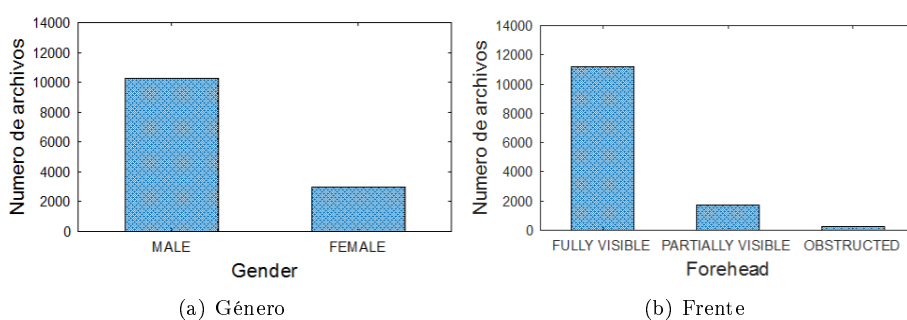


Figura 5.1: Histogramas Género y Frente

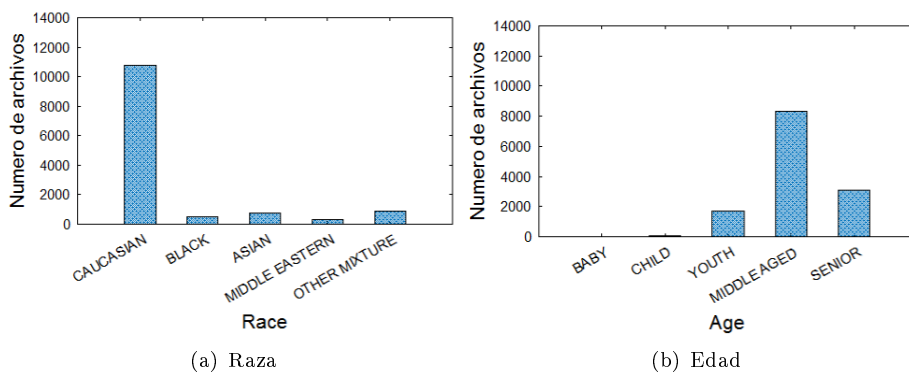


Figura 5.2: Histogramas Raza y Edad

GENDER		
	Male	Female
Num. imágenes	10264	2969
Porcentaje (%)	78	22

Tabla 5.1: Distribución de Soft Biometrics: Género

AGE					
	Baby	Child	Youth	Middle Aged	Senior
Num. imágenes	11	62	1708	8329	3123
Porcentaje (%)	<1	<1	13	63	24

Tabla 5.2: Distribución de Soft Biometrics: Edad

RACE					
	Caucasian	Black	Asian	Middle Eastern	Other Mixture
Num. imágenes	10795	505	733	318	882
Porcentaje (%)	81.5	4	5.5	2	7

Tabla 5.3: Distribución de Soft Biometrics: Raza

FOREHEAD			
	Fully Visible	Partially Visible	Obstructed
Num. imágenes	11213	1758	262
Porcentaje (%)	84.7	13.3	2

Tabla 5.4: Distribución de Soft Biometrics: Frente

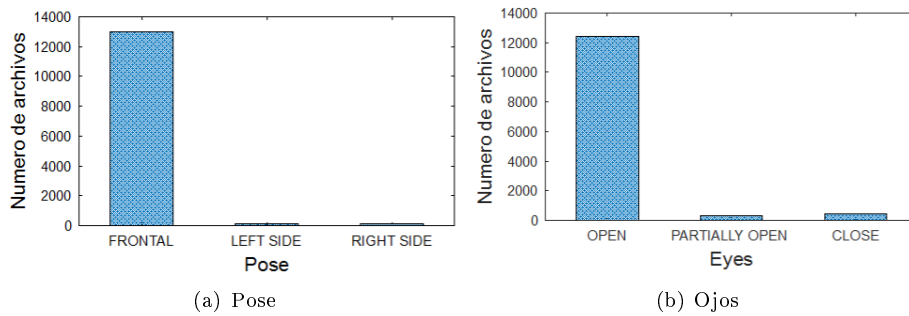


Figura 5.3: Histogramas Pose y Ojos

POSE			
	Frontal	Left Side	Right Side
Num. imágenes	13006	124	103
Porcentaje (%)	98	0.97	0.78

Tabla 5.5: Distribución de Soft Biometrics: Posición

EYES			
	Open	Partially Open	Close
Num. imágenes	12445	344	444
Porcentaje (%)	94	2.6	3.4

Tabla 5.6: Distribución de Soft Biometrics: Ojos

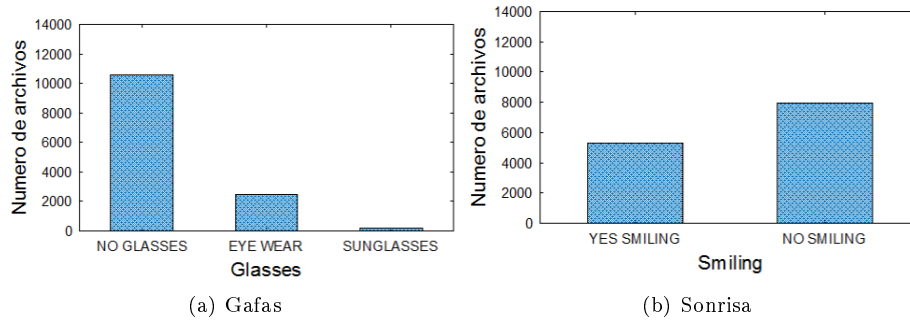


Figura 5.4: Histogramas Gafas y Sonrisa

GLASSES			
	No Glasses	Eye Wear	Sunglasses
Num. imágenes	10585	2465	183
Porcentaje (%)	80	18.6	1.4

Tabla 5.7: Distribución de Soft Biometrics: Gafas

SMILING		
	Yes	No
Num. imágenes	5309	7924
Porcentaje (%)	40.1	59.9

Tabla 5.8: Distribución de Soft Biometrics: Sonrisa

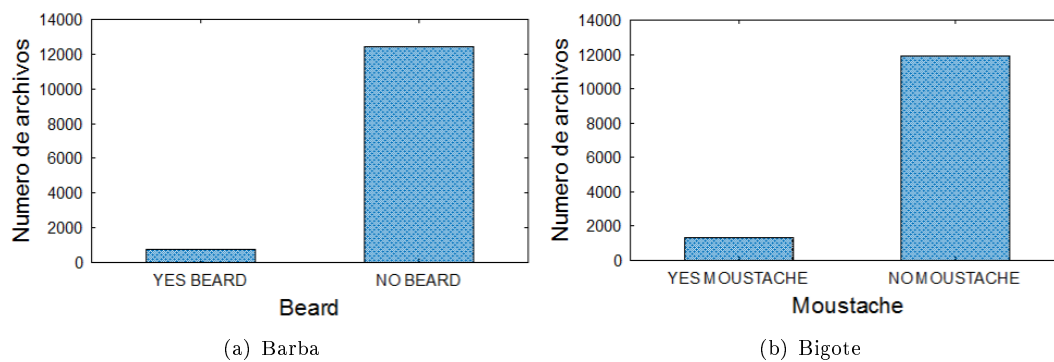


Figura 5.5: Histogramas Barba y Bigote

BEARD		
	Yes	No
Num. imágenes	786	12447
Porcentaje (%)	6	94

Tabla 5.9: Distribución de Soft Biometrics: Barba

MOUSTACHE		
	Yes	No
Num. imágenes	1325	11908
Porcentaje (%)	10	90

Tabla 5.10: Distribución de Soft Biometrics: Bigote

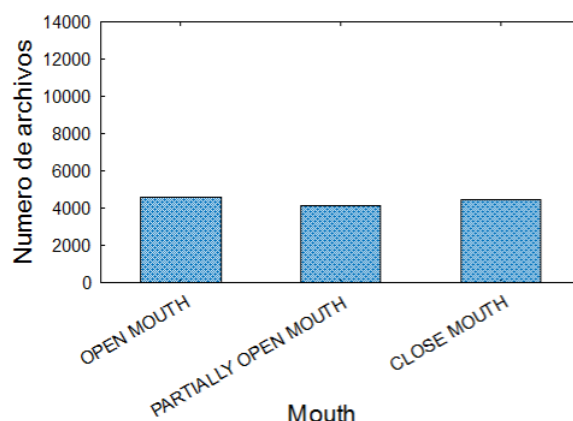


Figura 5.6: Histograma de Boca

MOUTH			
	Open Widely	Partially Open	Close
Num. imágenes	4604	4163	4465
Porcentaje (%)	34.8	31.5	33.7

Tabla 5.11: Distribución de Soft Biometrics: Boca

Una vez analizados todos los histogramas, se puede observar que la base de datos LFW tiene una predominancia de hombres frente a mujeres. En cuanto a la edad, destacan las personas de media edad, sin embargo el número de bebés y niños es muy reducido. Si se presta atención a la Raza, se observa que predominan las personas de raza caucásica frente a la raza negra, asiática, oriental o de otras razas. Atendiendo a la frente, la mayoría de las personas de la base de datos tienen la frente totalmente visible y muy pocos de ella la tiene obstruida. Así mismo, se puede observar que predominan los individuos con los ojos abiertos frente a aquellas personas que los tienen cerrados o parcialmente abiertos. Observando los histogramas de las gafas, destaca el número de personas que no llevan gafas frente a los que llevan gafas de ver o los que llevan gafas de sol. En cuanto a la barba y el bigote, predominan las personas que no tienen ninguna de estas dos características frente a aquellas personas que si las tienen. Sin embargo, observando la boca, el número de personas que la tienen abierta, parcialmente abierta o cerrada es similar.

Analizando la posición de los individuos de esta base de datos, casi todos ellos tienen una posición frontal, siendo el número de posición lateral muy reducido.

5.2. Correlación de atributos

Se desea estudiar si existe correlación entre algunas características para analizar el sistema más detalladamente. La correlación entre las variables indica la fuerza y la dirección de una relación lineal de proporcionalidad entre dichas variables estadísticas. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos valores (A) y (B), existe correlación si al aumentar los valores de (A) lo hacen también los de (B) y viceversa. Para calcular la correlación entre las características de este proyecto se ha elegido el coeficiente de correlación de Pearson que se puede expresar de la siguiente forma:

$$\text{Coef. Correlación Pearson} = \frac{x_d \cdot y_d}{\|x_d\| \cdot \|y_d\|}$$

Donde

$$x_d = [x_1 - \bar{x}, x_2 - \bar{x} \dots x_i - \bar{x}]^T$$

$$y_d = [y_1 - \bar{y}, y_2 - \bar{y} \dots y_i - \bar{y}]^T$$

$$\bar{x} = \frac{1}{l} \sqrt{\sum_{i=1}^l x_i}$$

$$\bar{y} = \frac{1}{l} \sqrt{\sum_{i=1}^l y_i}$$

Interpretación del coeficiente de correlación:

- Si $r = 1$, existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si $0 < r < 1$, existe una correlación positiva.
- Si $r = 0$, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.
- Si $-1 < r < 0$, existe una correlación negativa.
- Si $r = -1$, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.

A nivel de atributos, se ha definido una matriz con los 11 atributos elegidos para crear la interfaz y se ha utilizado la correlación de Pearson para realizar el estudio. En el caso a nivel de atributos, se pueden extraer menos conclusiones que realizando la correlación a nivel de instancias, ya que son rasgos más generales. Sin embargo, hay tres puntos que resaltar: En la diagonal se presenta una correlación positiva perfecta ya que comparamos una serie de atributos con ellos mismos. En el caso de la correlación entre barba y bigote, nos encontramos con una correlación positiva, pues es más probable que una persona que tenga bigote tenga barba que si no tiene bigote. Existe también una correlación entre la Edad y las Gafas, esto tiene sentido si se piensa que a mayor edad, se utilizan más las gafas. Se observa una correlación negativa entre el atributo Smiling y el Género.

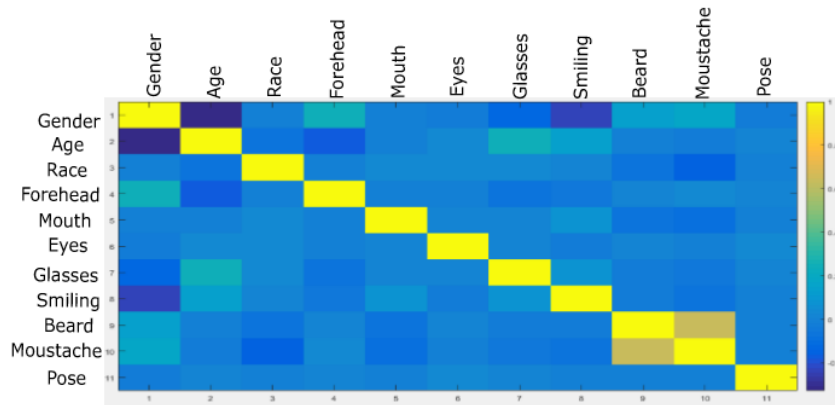


Figura 5.7: Correlación a nivel de atributos

A nivel de instancias, se ha definido una matriz con los 33 Soft Biometrics elegidos para crear la interfaz y se ha seguido el mismo proceso para calcular la correlación que la correlación a nivel de atributos. En el siguiente histograma, se puede observar que en la diagonal se presenta una correlación positiva perfecta ya que se está calculando la correlación de una serie de instancias con ellas mismas. Además, si se presta atención a los tonos amarillos, se puede ver como en el atributo de género, hombre y mujer presentan una correlación negativa perfecta, pues un individuo que sea hombre, no va a ser mujer. En el caso de la correlación entre tener bigote o barba y no tenerla, vemos una correlación negativa perfecta, ya que si tiene bigote no puede no tenerlo. Sin embargo, existe una correlación positiva, pero no perfecta, entre tener barba y tener bigote, pues si un individuo tiene bigote, lo más probable es que tenga barba y viceversa, pero no siempre se tiene por qué dar este caso. Se puede observar una correlación positiva entre los individuos que tienen la boca parcialmente abierta y que están sonriendo. Existe algo de correlación también entre las personas mayores y la instancia gafas de ver, pues probablemente las personas mayores necesiten más las gafas de ver que una persona joven o un bebé.

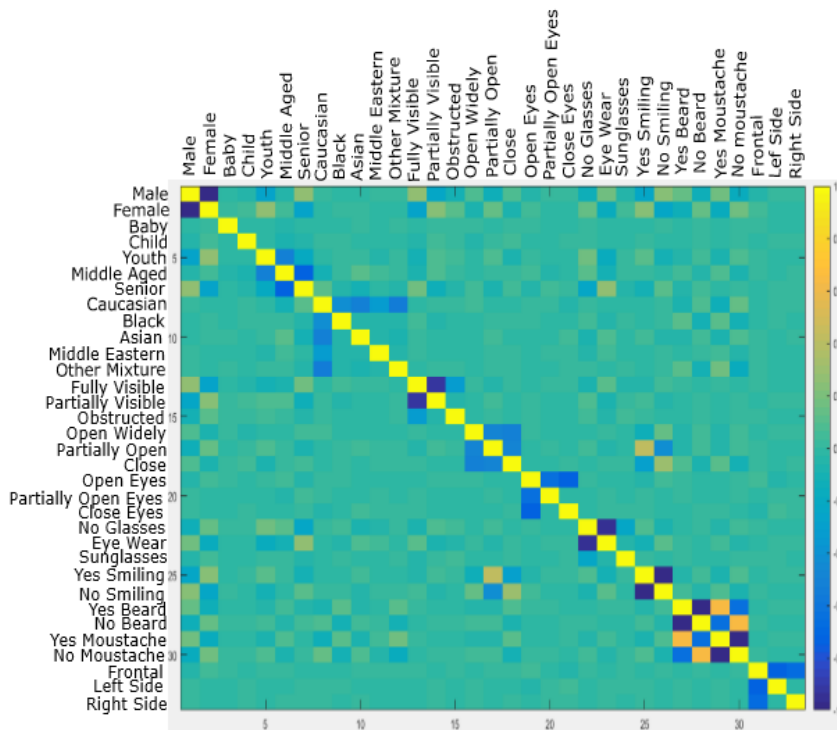


Figura 5.8: Correlación a nivel de instancias

6

Experimentos realizados y resultados

6.1. Medidas de rendimiento de sistemas biométricos

Para medir el rendimiento de un sistema de reconocimiento de usuarios mediante rasgos biométricos se utiliza un umbral preestablecido por experimentación, de forma que se pueden obtener dos resultados: aceptación o rechazo. Cuando una nueva muestra a verificar es introducida y comparada con una muestra registrada en la base de datos, en función de la puntuación resultante de la comparación, la muestra será reconocida como genuina siempre que dicha puntuación supere el umbral establecido. En relación a estas consideraciones, los posibles fallos del sistema serían aceptar como genuina una imagen que corresponde con otro usuario (FA-Falsa Aceptación) o bien establecer como falsa una imagen original de un usuario (FR- Falso Rechazo). Dado un considerable número de muestras genuinas e impostoras, así como un conjunto de pruebas de verificación, se pueden establecer las Tasas de Falsa Aceptación (FAR) y Falso Rechazo (FRR) para diferentes umbrales de decisión, permitiendo evaluar el rendimiento del sistema de reconocimiento biométrico. Con el fin de comparar el rendimiento entre diferentes sistemas de verificación, se suele utilizar como medida el punto en el cual la FAR y la FRR son iguales, es decir, la EER (Equal Error Rate), y que sería el punto de trabajo del sistema. Para establecer el punto de trabajo del sistema se utiliza la representación mediante curvas DET, en donde cada punto de la gráfica define un valor de FAR y de FRR, permitiendo la evaluación y comparación entre diferentes sistemas sin necesidad de tener en consideración múltiples curvas.

En la Figura 6.1 se puede ver una gráfica común de FAR, FRR y EER.

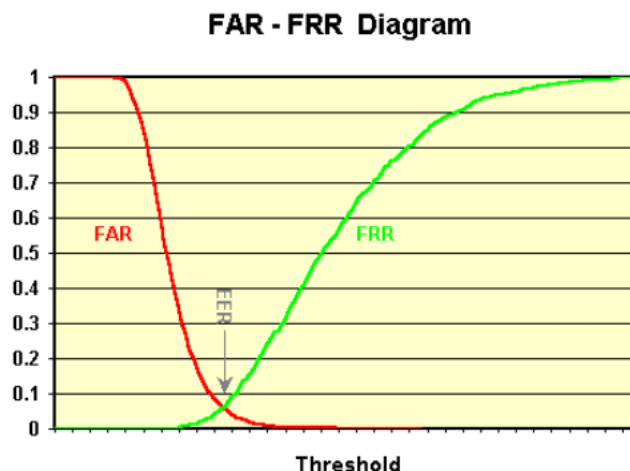


Figura 6.1: Ejemplo FAR, FRR y EER

Dado que los experimentos se han realizado diferenciando entre Soft Biometrics y Atributos, se puede citar una definición para cada uno de ellos. Soft Biometrics son aquellas características que presenta un individuo y que permanecen en el tiempo, es decir, que no cambian. Por ejemplo el género o la raza. Atributos son aquellas características que no son estables en el tiempo, por ejemplo la sonrisa (en un momento dado puedes estar feliz y en otro triste), la boca abierta o cerrada, etc. Sin embargo, hay ciertos rasgos de una persona que pueden ser considerados Soft Biometrics o atributos, dependiendo si permanecen siempre en ellas o no. Por ejemplo, una persona que siempre lleve bigote, será considerado como Soft Biometric, sin embargo, si una persona lleva bigote y al día siguiente no, será considerado atributo.

6.1.1. Basado en Soft Biometrics

Para obtener el rendimiento del sistema, se sigue el protocolo experimental estándar definido dentro de la evaluación de la LFW. Para la fase de evaluación, se definen 10 conjuntos (folds), con dos archivos cada uno: uno con parejas de imágenes pertenecientes a mismos individuos (llamado test same para obtener los scores genuinos del sistema), y otro con parejas de imágenes pertenecientes a diferentes individuos (llamado test diff para obtener los scores impostores). Cada archivo de test contiene 300 imágenes, ordenadas por parejas, seleccionadas de forma aleatoria, pero teniendo en cuenta que cada par de imágenes de test diff son diferentes entre ellas, mientras que los pares de imágenes de test same son imágenes correspondientes al mismo individuo. Una vez divididas las imágenes en estos ficheros de texto, se ha seleccionado como muestras genuinas las imágenes de las carpetas test same, es decir aquellas imágenes que pertenecen al mismo individuo, dejando como muestras impostoras aquellas imágenes recogidas en el fichero de texto test diff, es decir aquellos pares de imágenes que pertenecen a distintos individuos. Posteriormente se procede a calcular el número de muestras genuinas y el número de muestras impostoras, realizándolo tanto para una distancia Hamming como para una distancia Euclídea. Finalmente, se calcula el EER en base al número de muestras genuinas y muestras impostoras, calculando el EER para cada fold y luego calculando la media y su desviación típica.

Este experimento se realiza tanto para Soft Biometrics como para Atributos. En la Tabla 6.1 podemos observar que el rendimiento del sistema para Soft Biometrics es igual tanto para una distancia Hamming como para una distancia Euclídea.

	EER (EUCLIDEAN)	EER (HAMMING)
INSTANCIAS	26,600 \pm 2,260	26,600 \pm 2,260

Tabla 6.1: EER basado en Soft Biometrics

6.1.2. Basado en Atributos

El mismo experimento descrito anteriormente se realiza para el sistema basado en Atributos, observando que en este caso el EER es menor para una distancia Hamming que para una distancia Euclidea. Los resultados se pueden observar en la Tabla 6.2:

	EER (EUCLIDEAN)	EER (HAMMING)
ATRIBUTOS	$28,900 \pm 1,899$	$26,600 \pm 2,260$

Tabla 6.2: EER basado en atributos

6.1.3. Rendimiento individual

Una vez obtenido el rendimiento general del sistema tanto para Soft Biometrics como para Atributos, se ha realizado un rendimiento individual del sistema, en el que se ha calculado el EER para cada fold y para cada atributo definido en la Interfaz gráfica y posteriormente se procede a calcular la media de las 10 folds y su desviación típica. (Véase Tabla 6.3).

ATRIBUTO	EER
gender	$61,866 \pm 3,140$
age	$50,600 \pm 3,1419$
race	$64,333 \pm 3,910$
forehead	$68,200 \pm 2,535$
mouth	$55,133 \pm 3,248$
eyes	$88,900 \pm 3,201$
glasses	$68,833 \pm 2,626$
smiling	$52,866 \pm 3,433$
moustache	$86,266 \pm 2,216$
beard	$80,166 \pm 2,924$
pose	$96,833 \pm 1,157$

Tabla 6.3: EER para cada atributo

Una vez obtenido el rendimiento individual, se procede a calcular el EER de Soft Biometrics y de los atributos por separado. El resultado se puede ver en la tabla 6.4 :

C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	EER
Gender	Age	Race									$19,100 \pm 3,259$
Forehead	Mouth	Eyes	Glasses	Smiling	Moustache	Beard	Pose				$37,900 \pm 2,505$

Tabla 6.4: EER de Soft Biometrics y de atributos

Observando los resultados de la tabla anterior, se detecta que el conjunto de Soft Biometrics minimiza enormemente el EER comparado con el conjunto de atributos.

6.1.4. Selección de características a través de SFFS

El rendimiento de un sistema se degrada cuando los datos de entrenamiento son demasiados pequeños comparados con el número de dimensiones del vector de características. La utilización de técnicas de selección de características pretende disminuir la dimensionalidad del vector de características con el objetivo de mejorar el rendimiento del sistema.

Con el método SFFS (Sequential Forward Floating Search), algoritmo utilizado para extracción de características, dado un conjunto de F características, el primer paso es seleccionar aquella características más discriminadora x_i . Posteriormente, se prueban todas las combinaciones posibles entre esta característica x_i y el resto de características disponibles, seleccionando la mejor combinación $\{x_i, x_f\}$. Este

algoritmo se repite de manera iterativa añadiendo características hasta alcanzar el tamaño del vector de características deseado. La implementación de dicho algoritmo se puede ver en [7].

Para mejorar el rendimiento del sistema, el entrenamiento del algoritmo SFFS se ha realizado con la partición de development de la base de datos LFW, que es disjunta de los 10 folds de la fase de evaluación. En la tabla 6.6 se pueden ver los resultados obtenidos.

C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	EER
Age											45
Age	Race										22
Age	Race	Gender									14
Age	Race	Gender	Moustache								10
Age	Race	Gender	Moustache	Glasses							8
Age	Race	Gender	Moustache	Glasses	Beard						8
Age	Race	Gender	Moustache	Glasses	Beard	Pose					8
Age	Race	Gender	Moustache	Glasses	Beard	Forehead	Mouth				13
Age	Race	Gender	Moustache	Glasses	Beard	Pose	Forehead	Mouth			13
Age	Race	Gender	Moustache	Glasses	Beard	Forehead	Pose	Eyes	Smiling		19
Age	Race	Gender	Moustache	Glasses	Beard	Forehead	Mouth	Pose	Eyes	Smiling	23

Tabla 6.5: Rendimiento del sistema con datos de entrenamiento basado en el algoritmo de extracción de características SFFS

A continuación, se han obtenido resultados con el mismo conjunto de variables que ha resultado del algoritmo SFFS de los datos de entrenamiento, pero realizándolo para las 10 folds definidas anteriormente.

C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	EER
Age											$50,600 \pm 3,142$
Age	Race										$31,133 \pm 3,863$
Age	Race	Gender									$19,100 \pm 3,259$
Age	Race	Gender	Moustache								$16,033 \pm 3,093$
Age	Race	Gender	Moustache	Glasses							$11,967 \pm 2,598$
Age	Race	Gender	Moustache	Glasses	Beard						$11,90 \pm 2,183$
Age	Race	Gender	Moustache	Glasses	Beard	Pose					$14,133 \pm 1,5571$
Age	Race	Gender	Moustache	Glasses	Beard	Forehead	Mouth				$16,167 \pm 1,854$
Age	Race	Gender	Moustache	Glasses	Beard	Pose	Forehead	Mouth			$16,633 \pm 2,285$
Age	Race	Gender	Moustache	Glasses	Beard	Forehead	Pose	Eyes	Smiling		$18,50 \pm 2,2621$
Age	Race	Gender	Moustache	Glasses	Beard	Forehead	Mouth	Pose	Eyes	Smiling	$26,60 \pm 2,2597$

Tabla 6.6: Rendimiento del sistema con los 10 folds basado en el algoritmo de extracción de características SFFS

Finalmente se procede a aplicar el mismo algoritmo de extracción de características con las imágenes de la base de datos LFW y las características aplicadas en ella. A diferencia del anterior procedimiento, se han ido añadiendo las características con menor EER individual. Los resultados se pueden observar en la Tabla 6.7.

C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	EER
Age											$50,600 \pm 3,142$
Age	Smiling										$35,066 \pm 3,744$
Age	Smiling	Mouth									$44 \pm 2,977$
Age	Smiling	Mouth	Gender								$32,63 \pm 3,148$
Age	Smiling	Mouth	Gender	Race							$25,200 \pm 1,807$
Age	Smiling	Mouth	Gender	Race	Forehead						$29,400 \pm 2,614$
Age	Smiling	Mouth	Gender	Race	Forehead	Glasses					$32,20 \pm 2,466$
Age	Smiling	Mouth	Gender	Race	Forehead	Glasses	Beard				$30,43 \pm 2,336$
Age	Smiling	Mouth	Gender	Race	Forehead	Glasses	Beard	Moustache			$28,70 \pm 1,746$
Age	Smiling	Mouth	Gender	Race	Forehead	Glasses	Beard	Moustache	Eyes		$27 \pm 2,261$
Age	Smiling	Mouth	Gender	Race	Forehead	Glasses	Beard	Moustache	Eyes	Pose	$26,60 \pm 2,2597$

Tabla 6.7: Rendimiento del sistema ordenando el menor EER utilizando el algoritmo de extracción de características SFFS

Como se puede observar, el rendimiento del sistema mejora considerablemente cuando se extrae más de una característica. Esto tiene sentido, ya que cuantas más características se consideren, más información se tendrá acerca del individuo.

7

Fusión con un sistema basado en cara

Atendiendo a lo mencionado durante este Trabajo Fin de Grado, Soft Biometrics están pensados para ayudar a un sistema biométrico clásico a ser más fiable y a facilitar el reconocimiento de un individuo en situaciones en las que la imagen no se ve con buena calidad, como algunas veces ocurre con cámaras de videovigilancia.

Para demostrar que esta idea es cierta, se ha calculado el EER de un sistema basado en cara y el EER de algunas características definidas en la interfaz, de tal forma que al fusionarlo, se minimiza considerablemente el EER bajo estudio. En la tabla 7.2 se muestran los resultados utilizando algunas de las características definidas en la interfaz (Age, Race, Gender, Glasses y Moustache).

EER_CARA	EER_SOFT BIOMETRICS	EER_FUSION
11,228	25,959	9,123

Tabla 7.1: Fusión de un sistema basado en cara con todas las características de la interfaz

EER_CARA	EER_SOFT BIOMETRICS	EER_FUSION
11,228	11,579	7,719

Tabla 7.2: Fusión de un sistema basado en cara con las características seleccionadas en SFFS de la interfaz

8

Conclusiones y trabajo futuro

Este trabajo tenía como objetivo el etiquetado de Soft Biometrics de una base de datos (LFW) de gran importancia en el reconocimiento biométrico, con el fin de ayudar a mejorar los algoritmos automáticos de extracción de Soft Biometrics.

Con esto, se ha pretendido motivar el uso de Soft Biometrics como el género, edad, raza, frente, boca, ojos, gafas, sonrisa, barba, bigote y posición para complementar el uso de un sistema biométrico primario, como el sistema biométrico facial, para conseguir el objetivo principal que es la identificación del ser humano. En los experimentos de fusión con un sistema basado en cara se puede observar que se mejora considerablemente el sistema al mezclar el sistema biométrico clásico con estos Soft Biometrics.

Ha sido motivado por el inminente desarrollo tecnológico que hace imprescindible contar con una base de datos que proporcione la información necesaria para la identificación de un individuo.

Antes de empezar con la etapa de pruebas, se ha realizado un estudio de la correlación entre todas las características obtenidas y posteriormente un análisis estadístico de la base de datos utilizada. Como resultado final de la correlación, se han detectado varios tipos de dependencias entre algunas características, por ejemplo, la existencia de correlación negativa perfecta entre hombre y mujer o la correlación positiva entre barba y bigote, como era de esperar.

La investigación acerca de los Soft Biometric está aumentando cada vez más debido a la gran cantidad de aplicaciones que pueden beneficiarse de la extracción de estas características. Algunos ejemplos de estas aplicaciones incluyen la seguridad o la vigilancia.

Glosario de acrónimos

- **API:** Application Programming Interface
- **EER:** Equal Error Rate
- **FA:** False Acceptance
- **FAR:** False Acceptance Rate
- **FR:** False Rejection
- **FRR:** False Rejection Rate
- **DNI:** Documento Nacional de Identidad
- **SFFS:** Sequential Forward Floating Selection
- **UAM:** Universidad Autónoma de Madrid

Bibliografía

- [1] Anil K Jain, Sarat C Dass, and Karthik Nandakumar. Can soft biometric traits assist user recognition? In Defense and Security, pages 561–572. International Society for Optics and Photonics, 2004.
- [2] Antitza Dantcheva, Carmelo Velardo, Angela D’angelo, and Jean-Luc Dugelay. Bag of soft biometrics for person identification. Multimedia Tools and Applications, 51(2):739–777, 2011.
- [3] Daniel A Vaquero, Rogerio S Feris, Duan Tran, Lisa Brown, Arun Hampapur, and Matthew Turk. Attribute-based people search in surveillance environments. In Applications of Computer Vision (WACV), 2009 Workshop on, pages 1–8. IEEE, 2009.
- [4] JL Wayman. Large-scale civilian biometric systems-issues and feasibility. In Proceedings of Card Tech/Secur Tech ID, volume 732, 1997.
- [5] Nawaf Almudhahka, Mark Nixon, and Jonathon Hare. Human face identification via comparative soft biometrics. In IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), 2016.
- [6] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.
- [7] Rubén Tolosana Moranchel et al. Estudio de interoperabilidad en sistemas biométricos de firma manuscrita dinámica. 2014.
- [8] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? a survey on soft biometrics. IEEE Transactions on Information Forensics and Security, 11(3):441–467, 2016.
- [9] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in ‘Real-Life’ Images: detection, alignment, and recognition, 2008.
- [10] Anil K Jain, Sarat C Dass, and Karthik Nandakumar. Soft biometric traits for personal recognition systems. In Biometric Authentication, pages 731–738. Springer, 2004.
- [11] Saurabh Panjwani and Achintya Prakash. Crowdsourcing attacks on biometric systems. In Symposium On Usable Privacy and Security (SOUPS 2014), pages 257–269, 2014.
- [12] Daniel Reid, Sina Samangooei, Cunjian Chen, Mark Nixon, and Arun Ross. Soft biometrics for surveillance: an overview. Machine learning: theory and applications. Elsevier, pages 327–352, 2013.
- [13] Pedro Tome, Julian Fierrez, Ruben Vera-Rodriguez, and Mark S Nixon. Soft biometrics and their application in person recognition at a distance. Information Forensics and Security, IEEE Transactions on, 9(3):464–475, 2014.



Manual del programador

A.1. Acerca de GUIDE

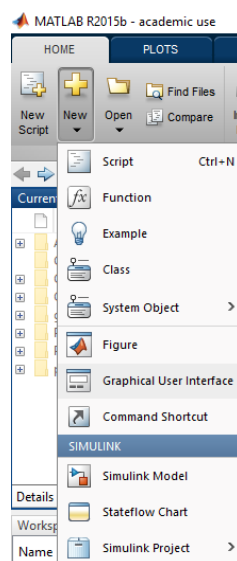
Para la realización de este Trabajo Fin de Grado, se ha utilizado la versión de Matlab R2015b.

Matlab permite crear GUIs de manera sencilla usando una herramienta llamada GUIDE(GUI Development Environment) La forma de implementar las GUI con Matlab es crear los objetos y definir las acciones que cada uno va a realizar. Al usar GUIDE se crearán dos archivos:

- Un archivo FIG: Representa los componentes que contenga la interfaz que se crea.
- Un archivo M: Contiene el código con las correspondencias de los botones de control de la interfaz.

Hay dos opciones para acceder a la herramienta GUIDE:

- Teclear `>> guide` en la ventana de comandos.
- A través de la opción HOME, haciendo clic en New y eligiendo la opción Graphical User Interface, (véase figura).



A continuación, se presenta el cuadro de diálogo de la Figura:

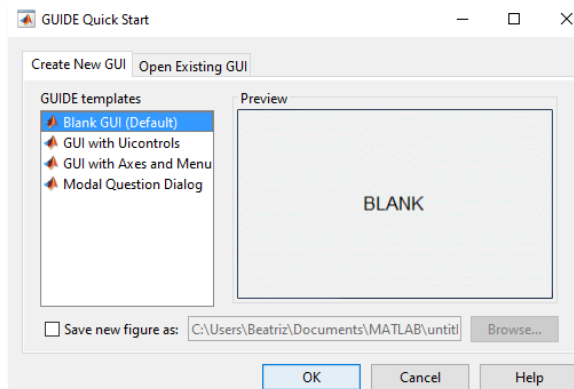


Figura A.1: Ventana de inicio GUI

Aparecen varias opciones:

- Blank GUI (Default)

La opción de interfaz gráfica de usuario en blanco, permite diseñar el programa que se desee.

- GUI with Uicontrols

Presenta un ejemplo de cálculo de masa, dada la densidad y el volumen. Este ejemplo se puede ejecutar y observar los valores obtenidos.

- GUI with Axes and Menu

Esta opción es otro ejemplo el cual contiene el menú File con las opciones Open, Print y Close. En el formulario tiene un Popup menu, un push button y un objeto Axes, se puede ejecutar el programa eligiendo alguna de las seis opciones que se encuentran en el menú desplegable y haciendo click en el botón de comando.

- Modal Question Dialog

Con esta opción se muestra en la pantalla un cuadro de diálogo común, el cual consta de una pequeña imagen, una etiqueta y dos botones Yes y No, dependiendo del botón que se presione, el GUI retorna el texto seleccionado (la cadena de caracteres Yes o No).

Si se elige la opción Blank GUI, Matlab mostrará una área de diseño similar a la figura. En la parte superior se encuentran los menús y opciones de GUIDE, en la parte izquierda se aprecian los diferentes controles y en la parte central el área de diseño donde se pondrán los controles a usar.

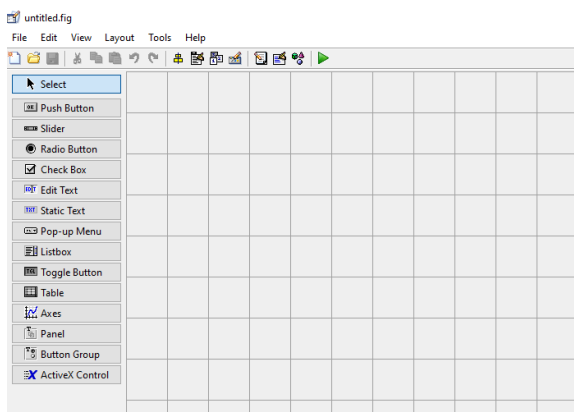


Figura A.2: Ventana de inicio GUI

A.2. Partes de GUIDE

La interfaz gráfica cuenta con las siguientes herramientas:

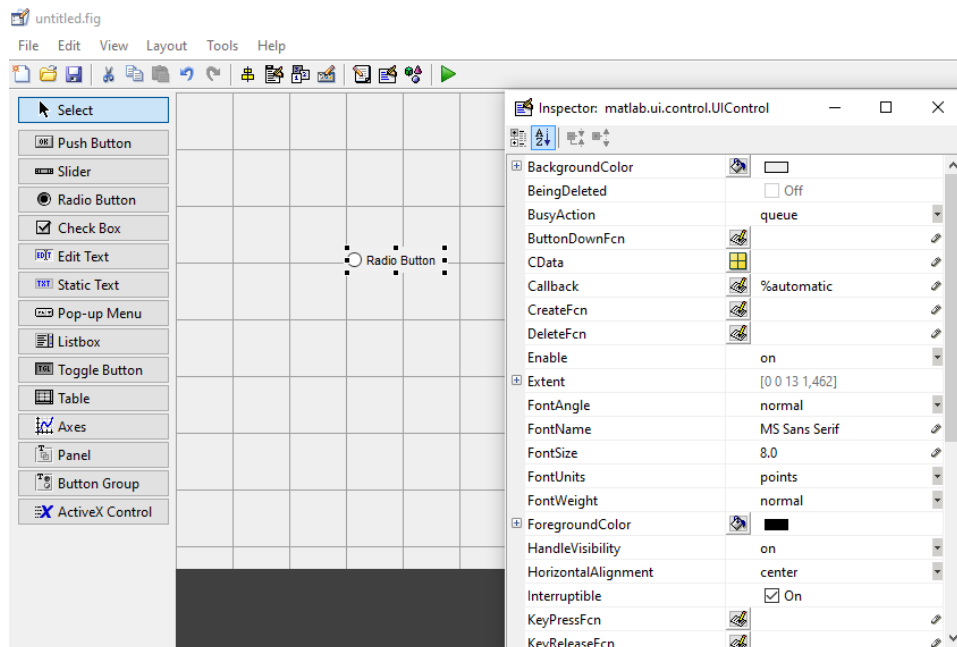
Herramienta	Aplicación
	Alinear objetos
	Editor de menú
	Editor de orden de etiqueta
	Editor del M-file
	Propiedades de objetos
	Navegador de objetos
	Grabar y ejecutar

Tabla A.1: Herramientas GUI Matlab

Herramienta	Aplicación
	Push button - Crea un botón.
	Radio Button - Crea un botón circular.
	Edit Text - Crea un campo de texto.
	Axes - Crea un área para gráficas.
	Static Text - Crea un letrero.
	Check Box - Permite hacer selecciones múltiples.
	Panel - Es la base donde se ponen todos los botones.

Tabla A.2: Otras Herramientas GUI Matlab

Una vez que se pone un botón, se puede cambiar sus propiedades en la opción de Property Inspector:



- Background Color - Cambia el color del fondo del control.
- CallBack - La propiedad más importante del control, ya que le dice al control qué hacer cuando este se active.

- Enable - Activa o desactiva un control.
- String - En el caso de botones, cajas de texto, texto estático; es el texto que muestra el control.
- Tag - Es otra de las propiedades mas importantes, ya que con ella es posible identificar el control.